

# Summary of Joint VCDE / Architecture Workspace Face-to-Face Session

ICR Workspace Teleconference

November 9, 2004

# Key Decisions at Face-to-Face # 2 (Chicago, IL – Nov 2004)

- ▶ The caGRID team within the Arch. workspace will embark on Phase II
  - Design & Prototyping based on Domain Workspace Use Cases
  - Particular Domain workspace projects will be use as vehicles for testing implementation strategies
  - These activities will yield
    - caBIG “Gold” Level Specification
    - “HOW-TO” – Guide for instantiating a caBIG Gold Grid Service
- ▶ Each data object and grid service must have a resolvable unique identifier.
  - Identifiers Subgroup will see look into the LSID specification to check if it provides the necessary framework for such an identifier system, but some additional specification would be needed
  - Provide a final design recommendation by December
- ▶ XML Schema alone may be insufficient to describe the complete semantics of a given data class or service
  - An accompanying metadata representation will be necessary
  - RDF/RDFS was proposed for this purpose
  - The caGRID team will evaluate RDF versus traditional XML and make a final design recommendation by December.

# Key Decisions at Face-to-Face # 2 (Chicago, IL – Nov 2004)

- ▶ Some level of formal data provenance will be necessary to describe data objects that are produced on the grid
  - Develop a minimal, general set of data elements for provenance that could apply to any type of data source.
  - Implementation will allow for extensions to the minimal set that may be needed for a particular sub-domain or measurement technology.
- ▶ A common query language that is standardized across all data grid services is needed.
  - The language will be expressible in XML
  - Should interact naturally with the object-oriented data models that will be queried and returned
  - A query Processor layer would be needed at a given data service provider's site
    - Translate the common query into a local query.
  - Need a set of language extensions, or possibly even a separate language, to enable semantic web, inference engine development, and reasoning across ontology structures. (Next Phase)

## Key Decisions at Face-to-Face # 2 (Chicago, IL – Nov 2004)

- ▶ HL7 does not satisfy all data messaging requirements for caBIG, but may apply to clinical data messaging
  - HL7 v3 messaging use cases need to be defined by Clinical Trials workspace and ICR
  - HL7 data types may be applicable outside of the clinical domain
  - HL7 V3 is not in wide use and V2 will need to be supported, but Workspace developers should be encouraged to consider V3 because of semantic richness
  - Final decisions on applicability of HL7 requires further study
- ▶ General agreement to caBIG data standards governance process with minor modifications based on feedback.
  - VCDE WS to develop detailed process for review, approval, and maintenance of data standards.
  - VCDE WS will consider procedures for notification of caBIG members regarding changes to data standards.
  - The VCDE WS will serve as the final review board for both NCI and caBIG-generated common data elements. The caDSR Context administrators outside of caBIG agreed, in principle, to grant the VCDE workspace review authority over their materials.

# Key Decisions at Face-to-Face # 2 (Chicago, IL – Nov 2004)

- ▶ A distributed model for vocabulary hosting is essential in order to avoid a single point of failure.
  - Agreed that multiple servers would be set up at nodes on the grid.
  - It was decided that vocabulary governance would be separated into two distinct topics:
    - Vocabulary Development
    - Vocabulary Deployment
  - The VCDE WS agreed that both topics are integrally related; however in order to move forward on decisions for each, they must be discussed separately.
  - Additionally, use cases from the Domain WS will drive decisions made about vocabulary development and deployment.